



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/519,640	12/28/2004	Sorin Georgescu	P16582-US1	9143
27045	7590	02/21/2008	EXAMINER	
ERICSSON INC. 6300 LEGACY DRIVE M/S EVR 1-C-11 PLANO, TX 75024			COLUCCI, MICHAEL C	
			ART UNIT	PAPER NUMBER
			2626	
			MAIL DATE	DELIVERY MODE
			02/21/2008	PAPER

**Please find below and/or attached an Office communication concerning this application or proceeding.**

The time period for reply, if any, is set in the attached communication.

**Office Action Summary**

Application No.

10/519,640

Applicant(s)

GEORGESCU, SORIN

Examiner

Michael C. Colucci

Art Unit

2626

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

- 1) ☐ Responsive to communication(s) filed on \_\_\_\_.
- 2a) ☒ This action is FINAL. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

- 4) ☒ Claim(s) 28-54 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 28-54 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_ are subject to restriction and/or election requirement.

**Application Papers**

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 28 December 2004 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.
- Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
- Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

- 12) ☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☒ Some \* c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
  2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_.
  3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☒ Information Disclosure Statement(s) (PTO/SB/08)
- Paper No(s)/Mail Date 12/28/2004.

- 4) ☐ Interview Summary (PTO-413)
- Paper No(s)/Mail Date. \_\_\_\_.
- 5) ☐ Notice of Informal Patent Application
- 6) ☐ Other: \_\_\_\_.

## DETAILED ACTION

### *Response to Arguments*

1. Applicant's arguments filed 12/21/2007 have been fully considered but they are not persuasive.

In response to arguments (pages 9-11):

Argument 1 (page 9 paragraph 3):

- "The Examiner stated that Lalitha discloses voice browsing, such that arbitrary web content can be accessed by voice commands without requiring conversion of the web content. The Applicant respectfully disagrees with this characterization"

Response to argument 1:

Examiner takes the position that Lalitha teaches a user agent that is a smaller version of the Web browser programs written for personal computers. These include programs such as MICROSOFT INTERNET EXPLORER and NETSCAPE NAVIGATOR. The user agent for the WAP-capable device is required to be smaller in size in order to fit in the memory of the device. The user agent must also download and render Web content equivalents (such as decks and cards) for a substantially smaller screen on the device than that used in a typical personal computer ([0026]).

Additionally, Lalitha teaches a block diagram of a typical wireless device (100) or mobile station. The device (100) is comprised of a microphone (105) for converting a voice signal to an electrical signal for transmission by the transmitter (103) and radiated over the antenna (109). The device user inputs information and operates the device by the

keypad (107). The keypad (107) can be used to input dual-tone multi-frequency (DTMF) responses. The display (106) shows the user what was input on the keypad (107) as well as information that was received by the receiver (104) ([0021], [0023] & Fig. 1 items 106 and 107).

Lalitha also teaches a multi-modal interface process of the present invention., where the user is accessing the Web site via a multi-modal wireless device. In this instance, multi-modal refers to the user agent supporting voice as well as data simultaneously for input and output on a user interface. The P3P preferences of this embodiment are set to multi-modal. Additionally, Lalitha teaches a WAP proxy that invokes the voice browsing Web service. In this embodiment, the user is accessing the Web site with a wireless device that has limited processing capabilities, such as a WAP-enabled device (Fig. 7 and [0069]).

Furthermore, Examiner takes the position that Kredo teaches what Lalitha fails to teach, and would therefore be obvious to combine teachings, such as an IM proxy server that interacts with an audio browser to communicate with the telephony user via a telephone network and act as a proxy on behalf of the telephony user for the IM server. (10) The audio browser effectively translates speech-to-text for messages directed to the on line IM user and translates text-to-speech for messages received from the on line -user and directed to the telephony user. Similarly, messages directed to the telephony user via a mobile terminal or the like and received by the IM server from the on line IM user are forwarded to the IM proxy server. The IM proxy server will process the message to form a text-based message ready for conversion to an audio format.

The processed message is sent to the audio browser, which converts the message to an audio format and delivers it to the mobile terminal (Kredo col 1 line 64 - col 2 line 24).

Kredo teaches that in operation; the audio browser will receive a message and convert audible commands within the message for processing by the IM proxy server. The IM proxy server will receive the command derived from the audio message and create an instant message based on the message meaning and any associated characteristics. The instant message is then delivered to the on line IM user via the IM server (Kredo col 2 line 25-31). The IM proxy server 26 will generate the necessary call dialog in a VoiceXML page and provide the page to the audio browser 28. The audio browser 28 will execute the call dialog to control communications with the telephony user A via the mobile terminal 20, as well as deliver audio to the mobile terminal, and receive audio making up the message commands from the telephony user A (Kredo col 5 lines 1-40)..

The audio browser 28 provides text converted from audio to the IM proxy server 26 in the form of requests for web pages, and the responding web pages may include the text to convert and send to the mobile terminal 20 in an audible format. The call dialog provided in the VoiceXML pages may facilitate numerous iterations, instructions, and commands to effectively control the audio browser 28 and the connection with the mobile terminal 20 (Kredo col 1 lines 5-25).

Argument 2 (page 9 paragraph 4, page 10 paragraph 3, page 11 paragraph 2, page 11 paragraph 5):

- "Lalitha does not disclose that the proxy server recognizes/extracts key elements, using predefined rules, to trigger voice browsing, such that arbitrary web content can be accessed by voice commands. Specifically, Lalitha version of "voice browsing" does not enable the actual browsing of the Web site."

Response to argument 2:

Examiner takes the position that Lalitha teaches Using this rules-based language, a user can express his or her preferences in a set of preference-rules called a rule set. The rule set is then used by a software agent to make automated or semi-automated decisions regarding the acceptability of machine-readable privacy policies from P3P enabled Web sites ([0006]). Lalitha teaches that the user has set the preferences in his P3P user agent, such as through the APPEL rules, regarding conditions when he/she should be notified about the site's privacy policies. Automatic retrieval and processing of the XML policy then takes place. When the condition is triggered, the P3P user agent retrieves the natural language version of the privacy policy either automatically or at the explicit request of the user. In the browsing mode, the user wishes to access a Web site and retrieve a Web page and content or application. The user may or may not have visited the site previously. ([0043]).

Furthermore, Examiner takes the position that Kredo teaches what Lalitha fails to teach, and would therefore be obvious to combine teachings, such as the audio browser 28 provides text-to-speech and speech-to-text conversion to facilitate communications between the IM proxy server 26 and the mobile terminal 20. In addition to translating

text, the IM proxy server 26 may recognize commands and implement the commands.  
A short message service (SMS) gateway 30 or like system may be used to send alerts,  
instructions, or the like to the mobile terminal 20 outside of the IM services (Kredo col 4  
lines 1-26).

Kredo teaches that in operation; the audio browser will receive a message and  
convert audible commands within the message for processing by the IM proxy server.  
The IM proxy server will receive the command derived from the audio message and  
create an instant message based on the message meaning and any associated  
characteristics. The instant message is then delivered to the on line IM user via the IM  
server (Kredo col 2 line 25-31). The IM proxy server 26 will generate the necessary call  
dialog in a VoiceXML page and provide the page to the audio browser 28. The audio  
browser 28 will execute the call dialog to control communications with the telephony  
user A via the mobile terminal 20, as well as deliver audio to the mobile terminal, and  
receive audio making up the message commands from the telephony user A (Kredo col  
5 lines 1-40).

Argument 3 (page 9 paragraph 5, page 10 paragraph 4):

- "In addition, Kredo does not make up the missing elements. Kredo teaches speech recognition and the recognition of predefined words and phrases. Kredo does not teach or suggest using voice commands to perform actual voice browsing of a Web site."

Response to argument 3:

Examiner takes the position that Lalitha teaches the user agent then processes the policy and may need to retrieve the natural language version based on the preferences or a user action (e.g., key depression, voice command). If so, the user agent requests the natural language version of the policy by issuing a HTTP command such as "Get Natural Language Policy &discuri&gt;"(311). As is well known in the art, the "discuri" parameter is the Universal Resource Locator (URL) at which the natural language policy resides. The Web site responds with the natural language version of the policy to the user agent (312) ([0047]).

Lalitha teaches communication between the proxy and the Web service. If the policy is sent as a whole to the Web service, the proxy should retrieve the same before invoking the Web service. The Web service supports functions such as the ability to perform text-to-speech conversion and/or speech recognition, generate VXML compatible Web pages, and/or traverse them ([0089] & Fig. 6 items 615, 625, and 630).

Furthermore, Examiner takes the position that Kredo teaches what Lalitha fails to teach, and would therefore be obvious to combine teachings, such as the audio browser 28 provides text-to-speech and speech-to-text conversion to facilitate communications between the IM proxy server 26 and the mobile terminal 20. In addition to translating text, the IM proxy server 26 may recognize commands and implement the commands. A short message service (SMS) gateway 30 or like system may be used to send alerts, instructions, or the like to the mobile terminal 20 outside of the IM services (Kredo col 4 lines 1-26).



Kredo teaches that in operation; the audio browser will receive a message and convert audible commands within the message for processing by the IM proxy server. The IM proxy server will receive the command derived from the audio message and create an instant message based on the message meaning and any associated characteristics. The instant message is then delivered to the on line IM user via the IM server (Kredo col 2 line 25-31). The IM proxy server 26 will generate the necessary call dialog in a VoiceXML page and provide the page to the audio browser 28. The audio browser 28 will execute the call dialog to control communications with the telephony user A via the mobile terminal 20, as well as deliver audio to the mobile terminal, and receive audio making up the message commands from the telephony user A (Kredo col 5 lines 1-40).

Additionally, it well known for VoiceXML to implement a voice browser utilizing an interactive voice response system, where a computer can detect voice or touch tone entries for use with IVR applications utilizing a call flow, which is comparable to that of a traditional HTML web pages without audio response.

### ***Claim Rejections - 35 USC § 103***

2. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

The factual inquiries set forth in Graham v. John Deere Co., 383 U.S. 1, 148 USPQ 459 (1966), that are applied for establishing a background for determining obviousness under 35 U.S.C. 103(a) are summarized as follows: (See MPEP Ch. 2141)

- a. Determining the scope and contents of the prior art;
- b. Ascertaining the differences between the prior art and the claims in issue;
- c. Resolving the level of ordinary skill in the pertinent art; and
- d. Evaluating evidence of secondary considerations for indicating obviousness or nonobviousness.

3. Claims 28-32, 34-35, 39, 46-47, 53 are rejected under 35 U.S.C. 103(a) as being unpatentable over Suryanarayana, Lalitha, USPGPUB 20030112791 (herein after Lalitha) in view of Kredo et al, US 6816578 B1 (herein after Kredo).

Re claims 28-29, 34, and 53, Lalitha teaches a system for allowing multi-modal access of content over a global data communications network using a mobile station (MS) with a user agent, a proxy server, and a telephony platform ([0021], [0023] & Fig. 1 items 106 and 107), wherein

said mobile station is a dual mode station supporting concurrent voice and data sessions (Lalitha [0069]).

said proxy server comprises an enhanced functionality for supporting voice browsing ([0085] & fig. 9)

said telephony platform comprises an Automatic Speech Recognizer (ASR) and is operative to convert text messages to speech (Lalitha [0089])

when the proxy server recognizes/extracts said key elements, using predefined rules ([0043]), it triggers voice browsing, such that arbitrary web content can be accessed by voice commands without requiring conversion of the web content. ([0021], [0023] & Fig. 1 items 106 and 107)

However Lalitha fails to teach key elements are predefined and indicated in the original web content (Kredo col 5 line 26-40);

Kredo teaches the recognition of key elements relevant to the proxy server. Kredo teaches that speech recognition technology is effective and reliable in recognizing pre-defined words and phrases permitting the formation of a limited vocabulary or language. Recognized words or phrases are construed to be key elements within web content.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention multi-modal access of content using a mobile station, user agent, proxy sever, and telephony platform implementing speech recognition, rules, text to speech conversion and voice browsing, where key elements are recognized, such as a command. The use of speech recognition with VoiceXML would allow for user interaction with or without the use of the depression of a keyboard/keypad, and instead use verbal commands.

NOTE: Lalitha in view of Kredo fail to disclose a hyperlink associated with web content. However examiner takes official notice that it is well known to have hyperlinks within web content as part of html. Lalitha discloses web servers providing web content such as html ([0038]).

Re claim 30, Lalitha fails to teach the system of claim 28, wherein the proxy server parses (Kredo col 1 line 49-63) an accessed web content with regard to said key elements (Kredo col 5 line 26-40);

Kredo teaches the recognition of key elements relevant to the proxy server. Kredo teaches that speech recognition technology is effective and reliable in recognizing pre-defined words and phrases permitting the formation of a limited vocabulary or language. Recognized words or phrases are construed to be key elements within web content.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention multi-modal access of content using a mobile station, user agent, proxy sever, and telephony platform implementing speech recognition, rules, text to speech conversion and voice browsing, where key elements are recognized, such as a command. The use of speech recognition with VoiceXML would allow for user interaction with or without the use of the depression of a keyboard/keypad, and instead use verbal commands.

Re claim 31, Lalitha teaches the system of claim 28, wherein the accessed web content is browsed by means of key strokes or mouse clicks ([0023] & fig. 1).

Re claim 32, Lalitha teaches the system of claim 28, wherein said system allows for voice-based access of any tag based content ([0026]).

NOTE: A tag is construed as a type of markup.

Re claim 35, Lalitha fails to teach the system of claim 28, wherein the proxy server interfaces with the Automatic Speech Recognizer which comprises a medium size vocabulary speech recognizer (Kredo col 5 line 26-40)

Kredo teaches speech recognition technology is effective and reliable in recognizing pre-defined words and phrases permitting the formation of a limited vocabulary or language. A medium size vocabulary is construed to be a limited vocabulary if the vocabulary is not recited to be full.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention a proxy server interfacing with an automatic speech recognizer having a medium size vocabulary. Using a vocabulary would allow for the storage of data pertaining to the natural rules of a language as well as user specific commands.

Re claim 39, Lalitha teaches the system of claim 28, wherein the proxy server forwards text prompts to a text-to-speech function in the telephony Platform ([0089]), wherein the text messages are converted to speech and forwarded to the user ([0089]) over the voice channel set up by the proxy server ([0088])

Re claim 46, Lalitha teaches the system of claim 28, wherein the a request for voice browsing includes at least a voice browsing session ID ([0081]) and MSISDN of the user station ([0040])

Re claim 47, Lalitha fails to teach the system of claim 46, wherein the a user authenticated by the proxy server, a voice channel is established, concurrent with a data session channel, between the ASR and the mobile station (Kredo col 5 line 55-63)

Kredo teaches a proxy server that identifies a caller and accesses the users profile that includes passwords, logins, and preferences for the service (Kredo col 5 line 41-54) and the proxy server also identifies a user by processing identification information. Authentication is construed as the confirming of the identify of a user.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention authentication by a proxy server and a voice channel established between an ASR and a mobile station. Having authentication in a voice system would allow for a user to access his/her personal files and settings to have unique voice commands for a specific individual.

**4. Claims 33, 37-38, 43 are rejected under 35 U.S.C. 103(a) as being unpatentable over Suryanarayana, Lalitha, USPGPUB 20030112791 (herein after Lalitha) in view of Kredo et al, US 6816578 B1 (herein after Kredo) and further in view of Rhie et al US 5953392 A (herein after Rhie).**

Re claim 33, Lalitha in view of Kredo fail to teach the system of claim 28, wherein the user of the mobile station uses a key element indicated in the web content to select a specific hyperlink (Rhie col 2 line 12-24)

Rhie teaches a system that converts the information content of a web page from text to speech (voice signals), signals the hyperlink selections of a web page in an audio manner, and allows selection of the hyperlinks.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention selecting a specific hyperlink indicated by key elements within web content. The use of hyperlinks allows for the linking of one page to another when browsing web content whether using Voicexml or non-voice xml pages.

Re claim 37, Lalitha teaches system according to claim 28, wherein the predefined rules for voice key element extractions are simple rules ([0043])

However, Lalitha fails to teach relating to selection of a unique keyword (Kredo col 5 line 26-40);

Kredo teaches the recognition of key elements relevant to the proxy server. Kredo teaches that speech recognition technology is effective and reliable in recognizing pre-defined words and phrases permitting the formation of a limited vocabulary or language. Recognized words or phrases are construed to be key elements within web content.

However, Lalitha in view of Kredo fails to teach in the name of a hyperlink (Rhie col 2 line 12-24)

Rhie teaches a system that converts the information content of a web page from text to speech (voice signals), signals the hyperlink selections of a web page in an audio manner, and allows selection of the hyperlinks.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention text to speech conversion and voice browsing, where key elements are recognized, such as a command. The use of speech recognition with VoiceXML would allow for user interaction with or without the use of the depression of a keyboard/keypad, and instead use verbal commands. Additionally, it would have been obvious selecting a specific hyperlink indicated by key elements within web content. The use of hyperlinks allows for the linking of one page to another when browsing web content whether using Voicexml or non-voice xml pages.

Re claim 38, Lalitha in view of Kredo fails to teach predefined rules for voice key element extraction are numeric rules numbering hyperlinks in said web content (Rhie col 1 line 46-60)

Rhie teaches a system that converts the information content of a web page from text to speech (voice signals), signals the hyperlink selections of a web page in an audio manner, and allows selection of the hyperlinks. Additionally, Rhie teaches that in order for the user to access a hyperlink on the web page, the first web page needs to be faxed back to the user with the hyperlinks numerically annotated for reference:

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention numeric rules numbering hyperlinks in web content. Numbering hyperlinks allows for a user friendly access of a hyperlink from a voice command, where a number is more readily available for recognition than the entire hyperlink phrase.



Re claim 43, Lalitha teaches the system of claim 28, wherein the a connection is established between the proxy server and the Automatic Speech Recognizer of the telephony platform ([0087]) for specifying and identifying a called application to be accessed ([0081]).

**5. Claim 36 is rejected under 35 U.S.C. 103(a) as being unpatentable over Suryanarayana, Lalitha, USPGPUB 20030112791 (herein after Lalitha) in view of Kredo et al, US 6816578 B1 (herein after Kredo) and further in view of Groner US 6507643 B1.**

Re claim 36, Lalitha teaches the system of claim 28, wherein the predefined rules for voice key element extraction ([0043])

However, Lalitha in view of Kredo fails to teach syntactic rules (Groner col 6 line 45-51)

Groner teaches a syntax-by-rule speech recognition procedure 144 to recognize predefined known categories of speech.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention elements extracted using predefined syntactic rules. Using the syntax of text would allow for the proper conversion of voice messages to data information prior to the transmission of information to browse a Voicexml application, where a language must have specific syntax rules to recognize a users particular language.

6. **Claims 40-42, 44-45, 48-52, 54 are rejected under 35 U.S.C. 103(a) as being unpatentable over Suryanarayana, Lalitha, USPGPUB 20030112791 (herein after Lalitha) in view of Kredo et al, US 6816578 B1 (herein after Kredo) and further in view of Gong et al US 7177814 B2 (herein after Gong).**

Re claim 40, "between the conventional browser in the user agent and the speech browser in the proxy server ([0087])

However, Lalitha in view of Kredo fails to teach a synchronization engine is provided (Gong col 9 line 33-39 & fig. 1)

Gong teaches a system for synchronizing multiple modes where a server can has synchronized control to allow communication between devices.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention synchronization between a browser in the user agent and a browser in the proxy server. Using synchronization would allow for the correct transmission and receiving of voice data exchanged over a network, where proper recognition of voice commands would be transmitted.

Re claim 41, Lalitha in view of Kredo fails to teach the system of claim 40, wherein the proxy server (Gong col 22 lines 3-13) comprises a pushing mechanism for making the MS user agent refresh indicated, fetched content (Gong col 4 line 12-14 & Fig. 3).

Gong teaches a server-push process for synchronizing a browser after a voice gateway requests a VXML page and sends a message indicating a corresponding HTML page and updating an HTML page.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention a proxy server with a push mechanism to refresh content. Using a push mechanism would allow for a system to operate on a server-client for a client using a web browser, such as Voicexml, where version of webpages can be properly updated and pushed through to the client.

Re claims 42 and 50, "a semaphore object is introduced into the content returned to the proxy server for indicating activation or not of content refresh (Gong col 9 line 33-39 & fig. 1).

Gong teaches a system for synchronizing multiple modes where a server can has synchronized control to allow communication between devices. A semaphore is construed as an object used for the allowance of synchronization and communication.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention synchronization between a browser in the user agent and a browser in the proxy server. Using synchronization/semaphore would allow for the correct transmission and receiving of voice data exchanged over a network, where proper recognition of voice commands would be transmitted.

Re claim 44, Lalitha teaches the system of claim 43, wherein the proxy server comprises a number of subscriber records, and in that for each subscriber for which voice browsing should be supported, means for indication of voice browsing activation ([0087] & [0089]),

insertion in accessed web content, and which ([0085] & fig. 9), when selected, provides for establishment of a voice channel between the ASR and the mobile station ([0087])

However, Lalitha in view of Kredo fails to teach optional key element for triggering voice browsing or optional hyperlink name (Gong col 5 line 52-63)

Gong teaches a subscribe system having separate devices, each including one gateway, can be synchronized by keeping track of the IP addresses and port numbers of the separate devices, or by having the devices subscribe to the same topic at a publish/subscribe system (Gong col 19 line 43-55). Gong teaches a web server determining a hypertext markup language HTML.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention obvious a proxy server with subscriber records capable of using commands or key elements such as a hyperlink to trigger voice browsing, where a connection between an ASR and a mobile station is established. The use of hyperlinks allows for the linking of one page to another when browsing web content whether using Voicexml or non-voice xml pages.

Re claim 45, "if voice browsing is activated, the access request is forwarded from the proxy server ([0071]) to the relevant Application Service Provider, which returns the requested content to the proxy server ([0035]), and in that said proxy server comprises parsing and analyzing ([0072]), before forwarding the content as modified to the mobile station ([0021])"

finding and indicating key elements (Kredo col 5 line 26-40);

Kredo teaches the recognition of key elements relevant to the proxy server.

Kredo teaches that speech recognition technology is effective and reliable in recognizing pre-defined words and phrases permitting the formation of a limited vocabulary or language. Recognized words or phrases are construed to be key elements within web content.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention multi-modal access of content using a mobile station, user agent, proxy sever, and telephony platform implementing speech recognition, rules, text to speech conversion and voice browsing, where key elements are recognized, such as a command. The use of speech recognition with VoiceXML would allow for user interaction with or without the use of the depression of a keyboard/keypad, and instead use verbal commands.

Re claim 48, "keywords as recognized in voice commands ([0047]) from the end user are provided to the proxy server ([0088]), and in that the proxy server comprises, for finding the relevant link on which to send a request to the Application Service Provider, and in that the requested content, upon reception in the proxy server, is parsed, analyzed and pushed to the user agent"

However, Lalitha fails to teach stored key elements (Kredo col 5 line 26-40);

Kredo teaches the recognition of key elements relevant to the proxy server. Kredo teaches that speech recognition technology is effective and reliable in recognizing pre-

defined words and phrases permitting the formation of a limited vocabulary or language. Recognized words or phrases are construed to be key elements within web content.

However, Lalitha in view of Kredo fails to teach matching means for matching recognized voice commands (Gong col 2 line 1-5)

A web service provider and an application service provider where data from the provider allows for parsing. Gong also discloses a parse process having a voice recognition phase to recognize a string or strings (Gong fig. 15). Gong also discloses spoken data related to input matched to stored data within a grammar. Gong also discloses a user requesting a new html page by clicking on a link with a browser and the browser sending the request to a synchronization controller (Gong col 16 line 11-26).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention a proxy server matching voice commands with stored data to find a relevant link to send a request to the service provider where parsing and pushing take place prior to being sent to a user agent. Additionally, it would have been obvious to one of ordinary skill in the art at the time of the invention multi-modal access of content using a mobile station, user agent, proxy sever, and telephony platform implementing speech recognition, rules, text to speech conversion and voice browsing, where key elements are recognized, such as a command. The use of speech recognition with VoiceXML would allow for user interaction with or without the use of the depression of a keyboard/keypad, and instead use verbal commands. Using a push mechanism would allow for a system to operate on a server-client for a client using a web browser, such

as Voicexml, where version of webpages can be properly updated and pushed through to the client.

Re claims 49 and 51, Lalitha in view of Kredo fail to teach the system according to claim 39, wherein for synchronization (Gong col 9 line 33-39 & fig. 1) between the user agent of the mobile station and the proxy server (Gong col 22 lines 3-13), a client semaphore (Gong col 9 line 33-39 & fig. 1) object is introduced, by the proxy server (Gong col 22 lines 3-13), into the original content of which the original copy is stored in said server, and activated when voice browsed content is to be pushed to be mobile station (Gong col 4 line 12-14 & Fig. 3).

Gong teaches a system for synchronizing multiple modes where a server can has synchronized control to allow communication between devices. A semaphore is construed as an object used for the allowance of synchronization and communication. Gong also teaches a server-push process for synchronizing a browser after a voice gateway requests a VXML page and sends a message indicating a corresponding HTML page and updating an HTML page.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention synchronization between a browser in the user agent and a browser in the proxy server. Using synchronization/semaphore would allow for the correct transmission and receiving of voice data exchanged over a network, where proper recognition of voice commands would be transmitted.

It would also have been obvious to one of ordinary skill in the art at the time of the invention a proxy server with a push mechanism to refresh content. Using a push mechanism would allow for a system to operate on a server-client for a client using a web browser, such as Voicexml, where version of webpages can be properly updated and pushed through to the client.

Re claim 52, "the client semaphore object is created using a WML script variable ([0036]), fetched from the proxy server, and, in the proxy server ([0034]), a first and a second version of said script is stored, the first version comprising

However, Lalitha in view of Kredo fails to teach a script for semaphore activation, the second version comprising a script indicating semaphore inactive (Gong col 9 line 33-39 & fig. 1)

Gong teaches a system for synchronizing multiple modes where a server can has synchronized control to allow communication between devices. A semaphore is construed as an object used for the allowance of synchronization and communication. Gong also teaches a server-push process for synchronizing a browser after a voice gateway requests a VXML page and sends a message indicating a corresponding HTML page and updating an HTML page. Additionally, Gong teaches an embedded JavaScript command in the refresh reply to the browser, where the JavaScript command instructs the browser to load a new html page (Gong col 13 line 27-37).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention synchronization between a browser in the user agent and a



browser in the proxy server. Using synchronization/semaphore would allow for the correct transmission and receiving of voice data exchanged over a network, where proper recognition of voice commands would be transmitted.

Re claim 54, Lalitha teaches method for providing concurrent multi-modal access of Internet content from a dual mode mobile station ([0069]), said method comprising the steps of:

- providing an enhanced functionality proxy server ([0074]) supporting voice browsing ([0085]);

- establishing a connection between the enhanced proxy server ([0074]) and a telephony platform ([0087]) with an Automatic Speech Register (ASR) ([0085]);

- determining if voice browsing is to be active and, if so, performing the steps of:

- setting up a voice channel between the mobile station and the Automatic Speech Register; forwarding a request to the concerned application service provider ([0047]);

- parsing content ([0072])

- sending the modified content to the mobile station ([0069]);

- opening a voice browsing session ([0085]);

- opening a voice channel concurrent with a data session channel ([0085]);

However Lalitha fails to teach defining key elements to use for voice browsing (Kredo col 5 line 26-40);

keywords recognized in a user voice command with predefined and selected keywords to establish which link to use for sending a get request to the relevant application service provider (Kredo col 5 line 26-40)

analyzing paragraphs in the content to find key elements (Kredo col 5 line 26-40);

Kredo teaches the recognition of key elements relevant to the proxy server. Kredo teaches that speech recognition technology is effective and reliable in recognizing predefined words and phrases permitting the formation of a limited vocabulary or language. Recognized words or phrases are construed to be key elements within web content.

However, Lalitha in view of Kredo fails to teach the modifying, in the enhanced proxy, content by changing tag attributes to make key elements identifiable to the user (Gong col 14 lines 33-44)

processing and pushing the content received from the application service provider to the user agent (Gong col 4 line 12-14 & Fig. 3).

matching, in the enhanced proxy server (Gong col 16 line 11-26):

Gong teaches a process 600, referred to as no-input tag, for use with the system 200 includes the web server 240 sending the voice gateway 285 a VXML page with a no-input tag embedded (610). Every VXML page may have a no-input markup tag (&lt;no input&gt;) that specifies code on the voice gateway 285 to run if the voice gateway 285 does not receive any user input for a specified amount of time. The URL of a JSP is embedded in the code, and the code tells the voice gateway 285 to issue a HTTP get command to retrieve the JSP. The same no-input tag is embedded in every

VXML page sent to the voice gateway 285 and, accordingly, the no-input tag specifies the same JSP each time.

Gong teaches a server-push process for synchronizing a browser after a voice gateway requests a VXML page and sends a message indicating a corresponding HTML page and updating an HTML page.

Gong also discloses spoken data related to input matched to stored data within a grammar. Gong also discloses a user requesting a new html page by clicking on a link with a browser and the browser sending the request to a synchronization controller.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention multi-modal access of content using a mobile station, user agent, proxy sever, and telephony platform implementing speech recognition, rules, text to speech conversion and voice browsing, where key elements are recognized, such as a command. The use of speech recognition with VoiceXML would allow for user interaction with or without the use of the depression of a keyboard/keypad, and instead use verbal commands.

It would have been obvious to one of ordinary skill in the art at the time of the invention a proxy server with a push mechanism to refresh content. Using a push mechanism would allow for a system to operate on a server-client for a client using a web browser, such as Voicexml, where version of webpages can be properly updated and pushed through to the client.

***Conclusion***

7. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire **THREE MONTHS** from the mailing date of this action. In the event a first reply is filed within **TWO MONTHS** of the mailing date of this final action and the advisory action is not mailed until after the end of the **THREE-MONTH** shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than **SIX MONTHS** from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Michael C. Colucci whose telephone number is (571)-270-1847. The examiner can normally be reached on 9:30 am - 6:00 pm, Monday-Friday.


If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571)-272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Application/Control Number:  
10/519,640  
Art Unit: 2626

Page 28

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

Michael Colucci Jr.  
Patent Examiner  
AU 2626  
(571)-270-1847  
[Michael.Colucci@uspto.gov](mailto:Michael.Colucci@uspto.gov)



**RICHEMOND DORVIL**  
**SUPERVISORY PATENT EXAMINER**